Census Data Extractor
for GIS
Release 1.0

This program was written and tested for use with ATLAS GIS, Arc/View and Maptitude within the Bureau of Census. It is recommended that you have at least 2 Mb of free hard disk space on drive "C" as the program will create a temporary work file in this directory. All temporary files are deleted automatically when the program is done. It is important to note that this program works on one county at a time. The program code is provided as a public service of the Census Bureau; no warranty related to its suitability is implied or provided. This program is in the public domain and may not be copyrighted nor sold for profit.

The program is designed to eliminate the cumbersome task of manually creating a 'GEOID' field to match to the census geography polygons created with Census TIGER/Line™ files and imported into a GIS. As a user however, you must be well versed on the content and format of the various data files the program will import from CD-ROM. Specifically; PL 94-171, STF-1A, STF-1B, and STF-3A.

It is important to note that this version of the program will import all fields from the original data file. Field names (or matrices) are left intact. For example; P0020001 is the identifier for a specific data element from one type of CD-ROM. The P0020001 field from STF-1A is a different data element than the P0020001 field found on STF-3A.

Therefore, the technical documentation found with each type of CD-ROM is critical to the proper use of this program.

Filename Conventions

In addition to understanding the content of the data files found on the various CD-ROMS, you must also understand the database and index file naming conventions. There are a few rules of thumb to follow:

1) Be sure to know the correct database (dbf) file before running the program.

2) Know the appropriate index (ndx) to use.

3) Always type in the database and index file names with the extensions (.dbf or ndx).

4) Choose an appropriate directory and filename for the resulting import file that the program will create. This is the file that you will import into your GIS.

Index Files

Filename conventions for index files are easy. All PL 94-171 CD-ROMs use the following convention:

PL9417xx (where xx is the two character state)

STF indexes are a little more complicated:

STF1ATxx
STF1BTxx
STF3ATxx

(where xx is the two character state abbreviation)

These index files are for state, county, tract level searches. There are other index files on the CD-ROM. They are used when a search is conducted by MCD (CCD) or place. Incidentally, this program was written to extract data for three levels of census geography **only;** Tract/BNA, Block Group and Block)

Database Files

As with the index files, the PL 94-1 71 database (dbf) files are simple to use:

PL9417xx (where xx is the state abbreviation)

The PL .dbf filename is the same as the index (ndx) filename. The STF database filenames on CD-ROM are again more difficult:

STF-1A

FILE	MATRIX POSITIONS
STF1A0xx	P0010001 TO P0100010
STF1A1xx	P0110001 TO P0120062
STF1A2xx	P0120063 TO P0120155
STF1A3xx	P0120156 TO P0120248
STF1A4xx	P0120249 TO P0130031
STF1A5xx	P0130032 TO P0190040
STF1A6xx	P0200001 TO P0350002
STF1A7xx	P0360001 TO H0200002
STF1A8xx	H0210001 TO H0400009
STF1A9xx	H0410001 TO H0550004

For example, if you wanted median housing value from the STF-1A CD-ROM, you would consult the technical documentation and find that the data are in matrix position H023B001 (the documentation is in error at this point, it lists the matrix code as H23B0001 while in fact it is H023B001) By checking the table shown above or from the documentation, we can see that this matrix position is found within file 'STF1A8xx'. Thus, this would be the file name used with the program in order to extract median housing value for import into your GIS. Remember that the entire range of matrices is imported and that the field you want is found somewhere within that list.

STF-1B

FILE	MATRIX POSITIONS
STF1BXxx	All data elements are within this single file Again, xx is the state abbreviation

STF-3A

FILE	MATRIX POSITIONS
STF300xx	Data ID records (No census data here)
STF301xx	P1 - P13
STF302xx	P14A - P14C
STF303xx	P14D - P14F
STF304xx	P14G - P14I
STF305xx	P14J - P17
STF306xx	P18 - P26
STF307xx	P27 - P32
STF308xx	P33 - P36
STF309xx	P37 - P51
STF310xx	P52 - P60
STF311xx	P61 - P65
STF312xx	P66 - P70
STF313xx	P71 - P76
STF314xx	P77 - P83
STF315xx	P84 - P86
STP316xx	P87A
STF317xx	P87B
STF318xx	P87C
STF319xx	P87D
STF320xx	P8 7E
STF321xx	P88 - P107A
STF322xx	P108 - P118
STF323xx	P119 - P121
STF324xx	P122 -123
STF325xx	P124A - P124B
STF326xx	P125 - P143
STF327xx	P144 - P170
	Hl - H9
STF328xx	H10 - H21

STF329xx	H22 - H33
STF330xx	H34 - H44
STF331xx	H45 - H52A
STF332xx	H53 - H59
STF333xx	H60 - H81
STF334xx	H82 - H92

USING THE PROGRAM

Load the program onto your hard disk. Run the program with the “state” command from the DOS prompt .

After a brief introductory statement, the program will display a list of geographic levels to choose from. Choices range from I thru 6. Choosing 6 will terminate the program and return you to DOS. Note that there are differences between geographic levels on the different CD-ROMS.

If you choose a geographic level that is not available on a given CD-ROM, you will receive an error message and will have to re-start the program. Once you have chosen a geographic level, the program prompts you for a dBASE path and filename. At this point, type in the correct path and filename for the data you want. For example, **'D:\STF303DC.DBF'** is a valid response. The program will attempt to find the file you have indicated before proceeding. If it does not find the file, it will ask you to re-type the path and filename.

Once the database filename is entered, the program will prompt you for an index file to use. Again, type in a full path and filename that is an appropriate index for the . dbf file you are using. Keep in mind that the drive letter should always be the same as the **.** dbf file if you're extracting from CD-ROM. Once more, the program will check to see that what you have entered is a valid filename. If it is not valid, you are given the opportunity to make a correction.

Finally, the program will prompt you for an final import path and filename. Ideally this will be in the directory where your TIGER polygons are**.** The filename should reflect what data you working with and which geographic hierarchy you've extracted from. For example, a filename might be; STF1BG or VALUBG, etc.. This will be the file that you can import into your GIS.

During the extraction process, the program will create a new field in the database file called 'GEOID'. This will be the match field to use when importing the data into your GIS. Also, every field from the original data file is extracted in its original form from the CD-ROM. That means that the field names will still have an alpha-numeric identifier such as P0230034 or H0010001 and so on. It is up to you to know what these mean. Refer to the specific CD-ROM technical documentation for assistance.

Here is a sample of the code in dBase that can easily be ported to another language.

```
SET COLOR TO W/+
CLEAR
@ 6,4 SAY " "
@ 3,1 SAY "ENTER THE LEVEL OF GEOGRAPHY TO EXTRACT....."
@ 4,1 SAY " "
SET COLOR TO G/+
@ 5,5 SAY "P.L. or STF TRACT  = 1         STF BG = 4"
@ 6,5 SAY "          P.L. BG  = 2   STF-1B BLOCK = 5"
@ 7,5 SAY "       P.L. BLOCK  = 3           QUIT = 6"
@ 8,1 SAY " "
SET COLOR TO R/+
ACCEPT 'WHAT IS YOUR CHOICE?(1 - 6):' TO GEO
DO CASE
CASE GEO = "1"
 STORE '11' TO WIDE
 STORE '140' TO SUML
CASE GEO = "2"
 STORE '12' TO WIDE
 STORE '740' TO SUML
CASE GEO = "3"
 STORE '15' TO WIDE
 STORE '750' TO SUML
CASE GEO = "4"
 STORE '12' TO WIDE
 STORE '150' TO SUML
CASE GEO = "5"
 STORE '15' TO WIDE
 STORE '100' TO SUML
CASE GEO = "6"
 QUIT
OTHERWISE
 CLEAR
 @ 4,4 SAY 'WRONG SELECTION..... TRY AGAIN'
 WAIT
 QUIT
ENDCASE
CLEAR
SET COLOR TO W/+
ACCEPT ' ENTER THE dBASE PATH AND FILENAME:' TO INNAME
DO WHILE .NOT. FILE('&INNAME+.DBF')
CLEAR
@ 5,5 SAY "WRONG FILE NAME!    TRY AGAIN...."
```

```
@ 6,5 SAY " "
ACCEPT ' ENTER THE dBASE PATH AND FILENAME:' TO INNAME
ENDDO
ACCEPT ' ENTER THE INDEX PATH AND FILENAME:' TO IND
DO WHILE .NOT. FILE('&IND+.NDX')
CLEAR
@ 5,5 SAY " GEEZ!, WRONG FILE AGAIN......."
@ 6,5 SAY " Check the Documentation and try again..."
ACCEPT ' ENTER THE INDEX PATH AND FILENAME:' TO IND
ENDDO
ACCEPT 'ENTER THE IMPORT PATH AND FILENAME:' TO OUT
STORE '&SUML' TO CODE
USE &INNAME INDEX &IND
SET TALK ON
FIND &CODE
COPY TO C:\TEMP WHILE SUMLEV = "&SUML"
@ 10,10 SAY "   FILE COPIED FROM CD-ROM"
SET COLOR TO R/*
@ 12,10 SAY "   STARTING RAPID EXTRACTION....."
SET COLOR TO W/+
USE C:\TEMP
COPY TO BASE STRUCTURE EXTENDED
USE BASE
REPLACE FIELD_NAME WITH "GEOID" FOR FIELD_NAME="SUMLEV"
REPLACE FIELD_LEN WITH &WIDE FOR FIELD_NAME="GEOID"
CREATE &OUT FROM BASE
USE &OUT
APPEND FROM C:\TEMP
IF GEO = "1"
 REPLACE ALL GEOID WITH STATEFP+CNTY+TRACTBNA
ELSE
 IF GEO = "2" .OR. GEO = "4"
   REPLACE ALL GEOID WITH STATEFP+CNTY+TRACTBNA+BLCKGR
  ELSE
   IF GEO = "3" .OR. GEO = "5"
    REPLACE ALL GEOID WITH STATEFP+CNTY+TRACTBNA+BLCK
   ENDIF
 ENDIF
ENDIF
@ 10,10 SAY "PROCESS COMPLETE"
CLOSE DATABASES
ERASE C:\TEMP.DBF
ERASE BASE.DBF
CLEAR
ACCEPT 'HAVE ANOTHER GO AT IT? (Y/N)' TO DONE
```

```
IF DONE = 'Y' .or. DONE = 'y'
 DO STATE
ELSE
QUIT
ENDIF
```